

# HW1 Regression Report

113024510 廖振宇

## 1. Executive Summary 摘要

- 以純 `numpy` 實作的五層前饋網路 ( [d, 8, 8, 4, 1] ) 成功預測建築暖氣負載 ( Heating Load )。
- 在能源效率資料集 (768 筆) 上，最佳模型於 **第 11,604 個 epoch** 取得 `Validation RMS = 0.376`，最終測試誤差為 `Test RMS = 0.460`、`MAE = 0.334`。
- `Permutation Importance` 顯示 `roof_area`，`wall_area`，`relative_compactness` 為最關鍵特徵；僅保留單一或少數特徵時效能劇烈下降，證明多特徵整合的重要性。

## 2. Data & Preprocessing 資料與前處理

- 資料來源 Dataset**：HW1/datasets/2025\_energy\_efficiency\_data.csv。隨機打散後切分為 Train 461 / Val 115 / Test 192 (60/15/25)。
- 特徵工程 Feature Engineering**：
  - 六個連續欄位 ( `relative_compactness` 等) 以訓練集均值與標準差進行 z-score 標準化。
  - `orientation` (4 類) 與 `glazing_area_distribution` (6 類) 透過 one-hot encoding 加入模型，並維持固定欄位順序。
  - 目標變數 `heating_load` 保留原始尺度。
- 資料檔案產出 Artifacts**：
  - `artifacts/regression/training_history.csv`：每個 epoch 的 Train/Val MSE、RMS。
  - `results/regression_summary.json`：Train/Val/Test 的最終指標。

## 3. Model & Training Setup 模型與訓練設定

- 網路架構 Architecture**：[input\_dim, 8, 8, 4, 1]，隱藏層採 ReLU、輸出層線性。權重以 Xavier Uniform 初始化。
- 超參數 Hyperparameters**：

參數 Parameter	值 Value
Learning Rate	0.01
Epoch 上限	30,000
Mini-batch Size	32
Gradient Clip	$\pm 1.0$ (element-wise)
Early Stopping Patience	5,000 (監控 Validation RMS)
Random Seed	29
Gradient Check	Central difference, $\epsilon=1e-5$ , 抽樣 10 個權重 + 各層第 1 個 bias ; 最大相對誤差 $< 1e-3$

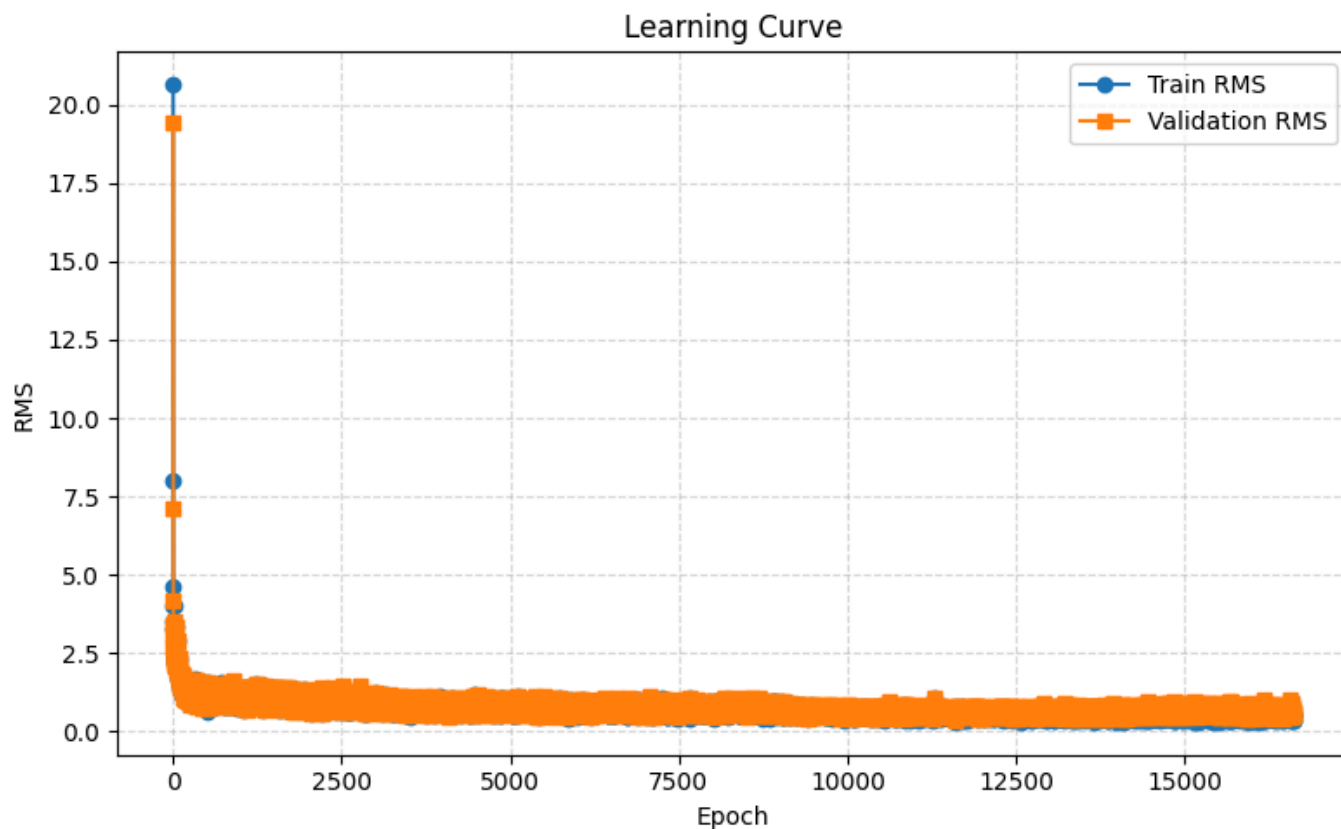
- **訓練流程 Highlights :**
  - 每個 epoch 重新打散 batch 順序 ( seed + epoch ) 。
  - 以完整 Train/Val 集計算 RMS 供早停監控。
  - 儲存最佳驗證表現時的參數 ( artifacts/regression/best\_model.pkl ) 。

## 4. Learning Dynamics 學習過程

- **最佳驗證點 Best Epoch :** 11,604

指標 Metric	Train	Val
MSE	0.0941	0.1414
RMS	0.3067	0.3760

- **學習曲線 Learning Curve :**



- 初期 ( $\leq 100$  epoch) RMS 急速下降，顯示神經網路快速捕捉主要模式。
- 之後平緩且有震盪，早停於 11.6k epoch 附近保留最佳驗證性能。
- 大量 epoch 後訓練 RMS 繼續下降但驗證 RMS 未再改善，顯示正規化不足但早停有效避免過度擬合。

## 5. Performance Evaluation 效能評估

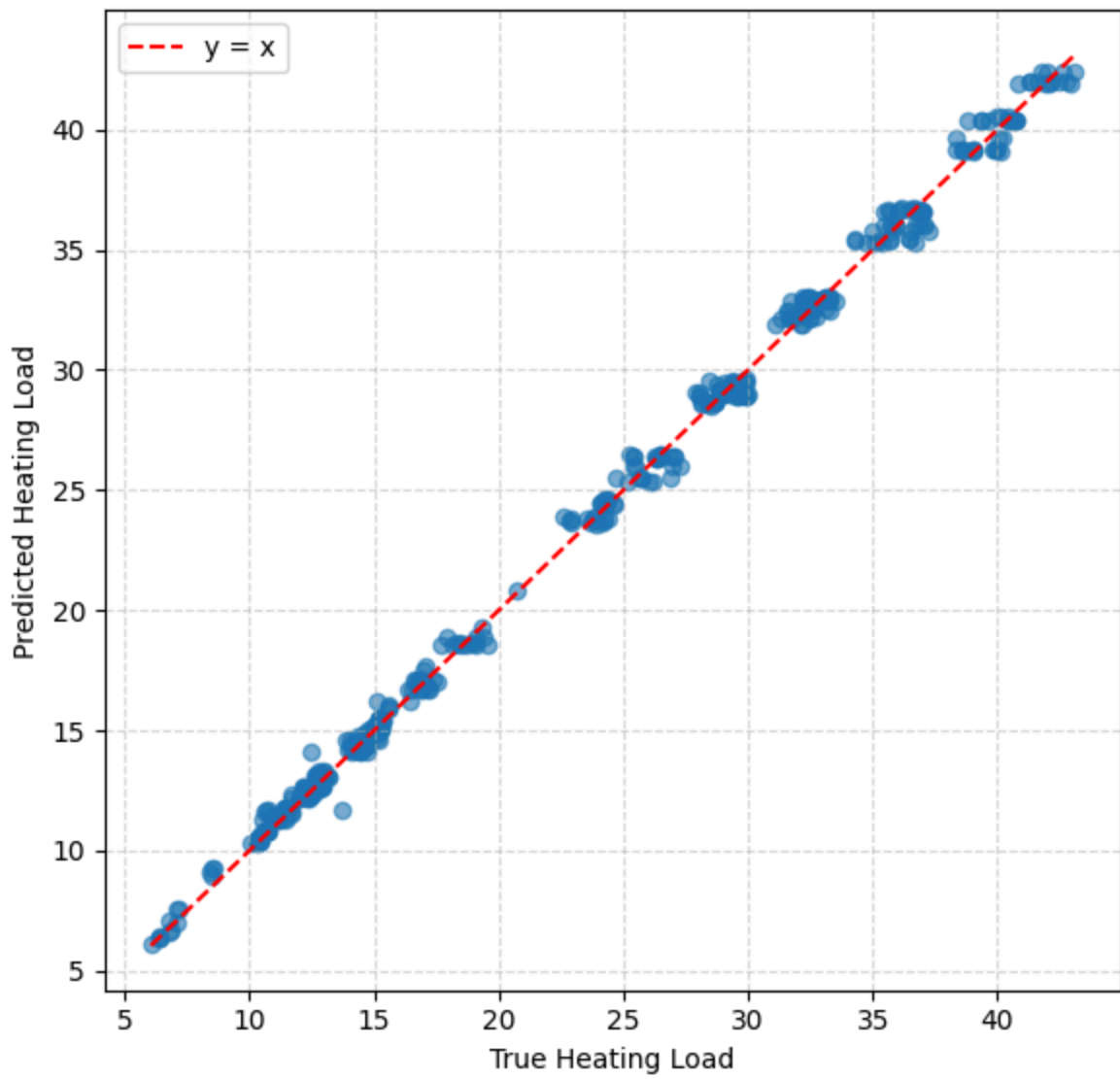
- 整體指標 Overall Metrics：

Split	MSE	RMS	MAE
Train	0.0941	0.3067	0.2310
Val	0.1414	0.3760	0.2993
Test	0.2114	<b>0.4598</b>	0.3341

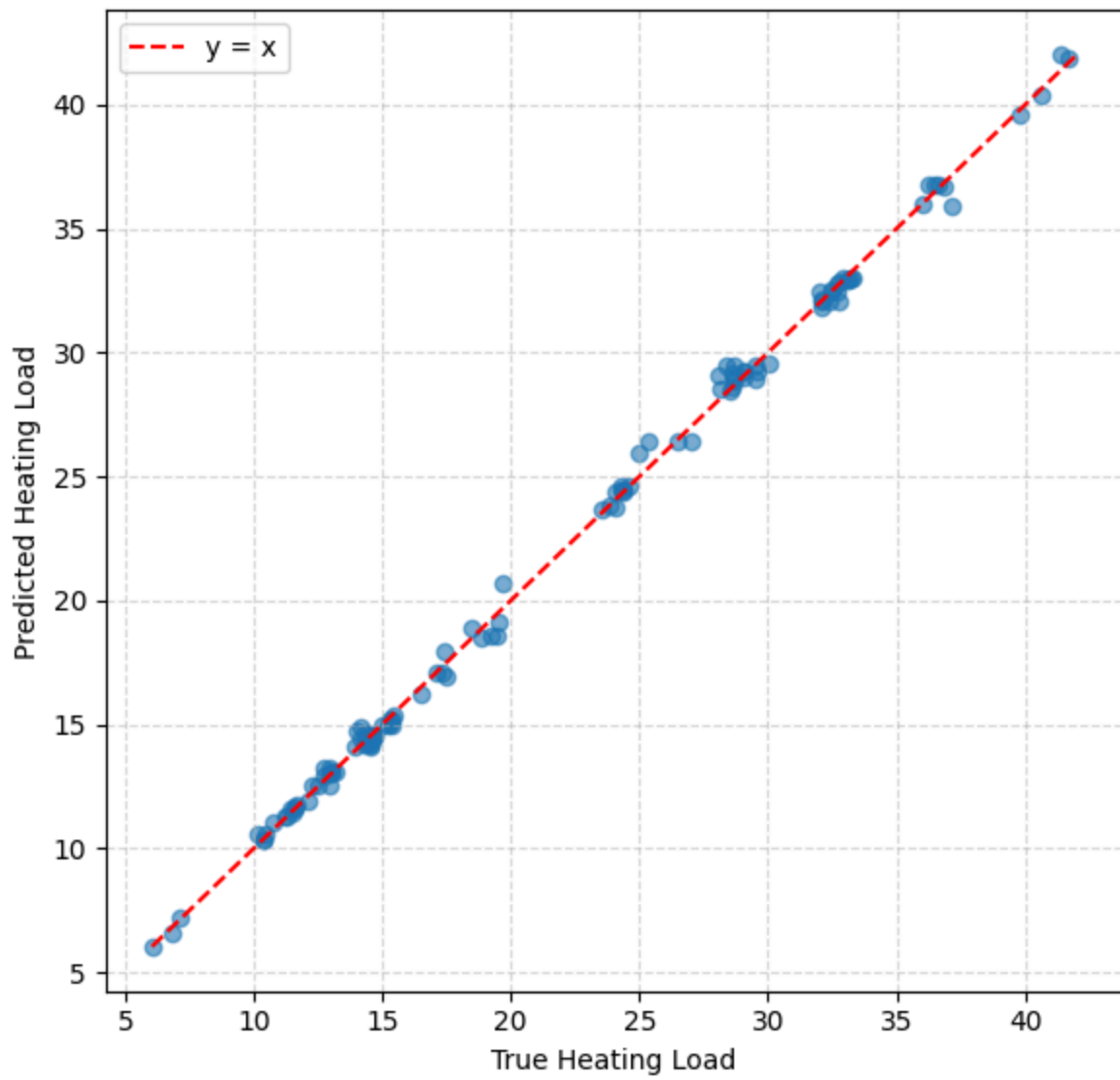
- 預測可視化 Prediction Visuals：

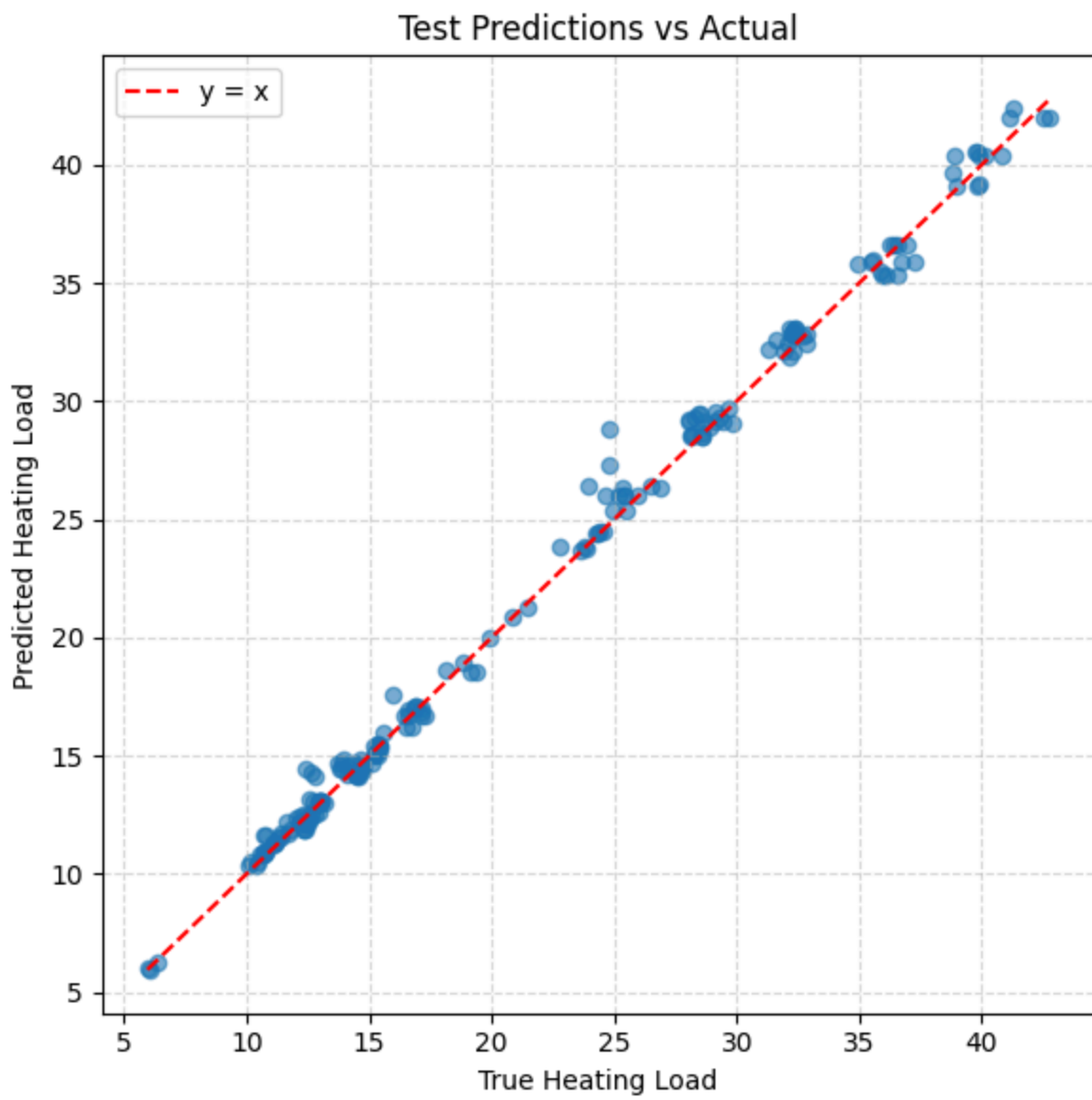
- Train / Val / Test  $y_{true}$  vs  $y_{pred}$ ：

Train Predictions vs Actual

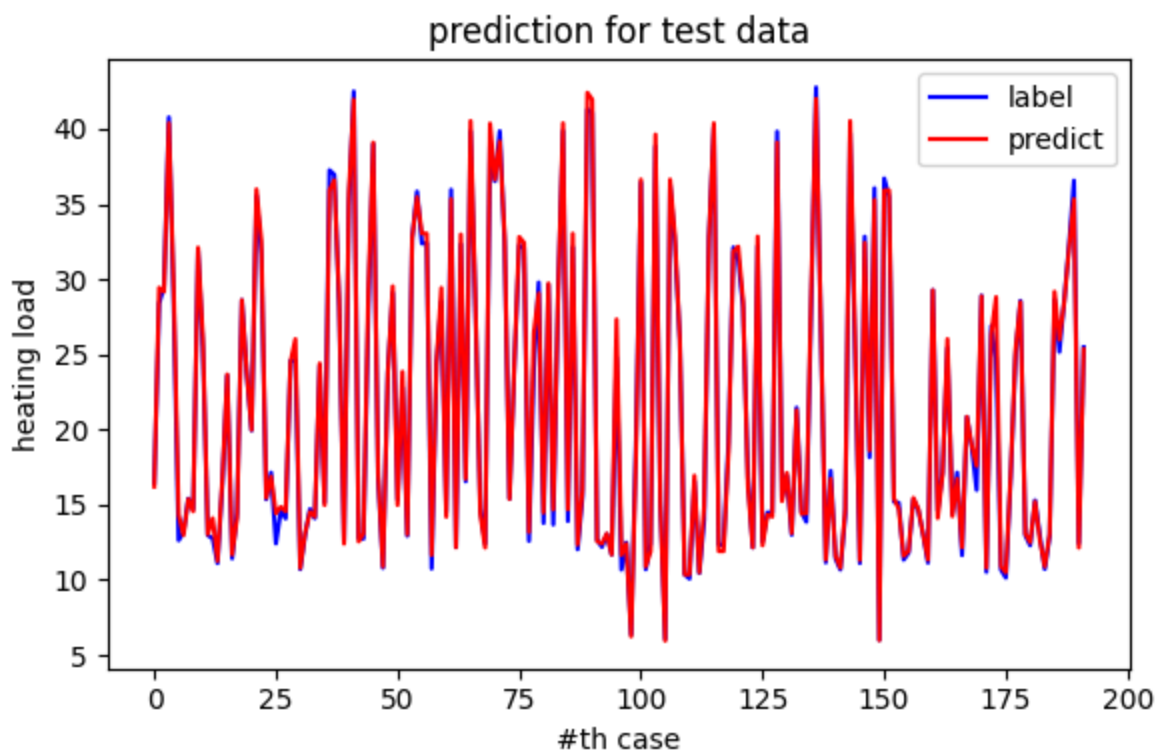
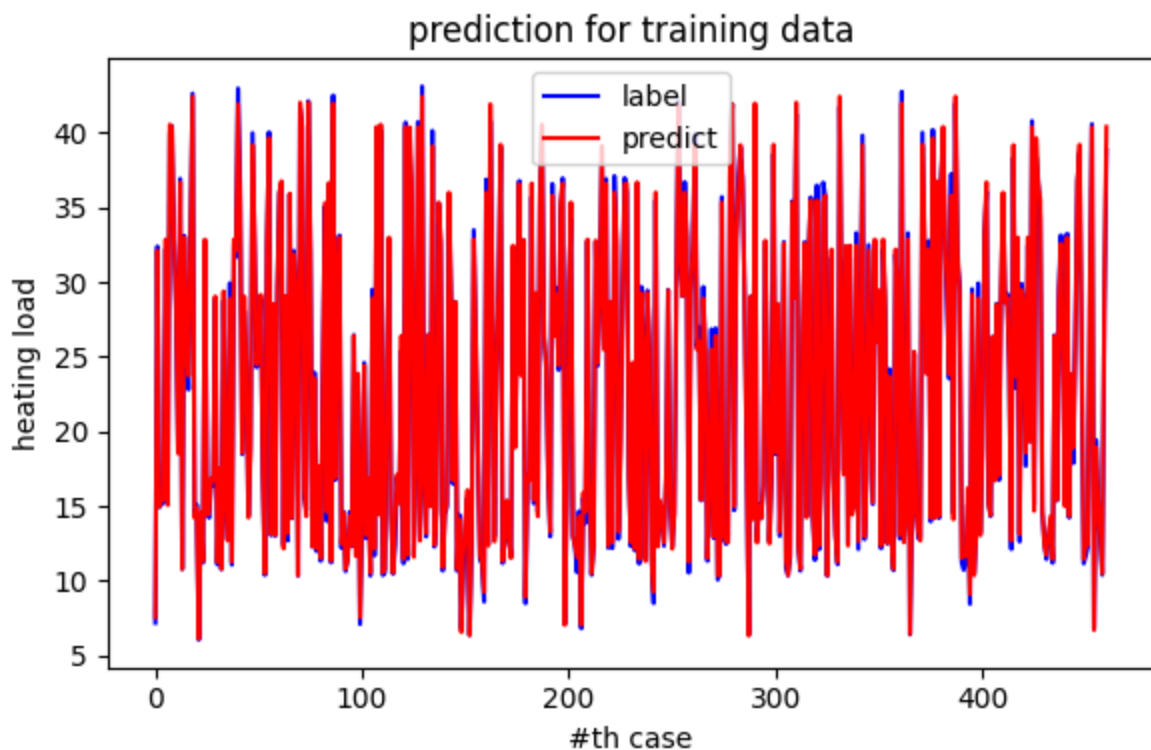


Val Predictions vs Actual





- Train/Test 預測與真值折線：



- **觀察 Observations :**

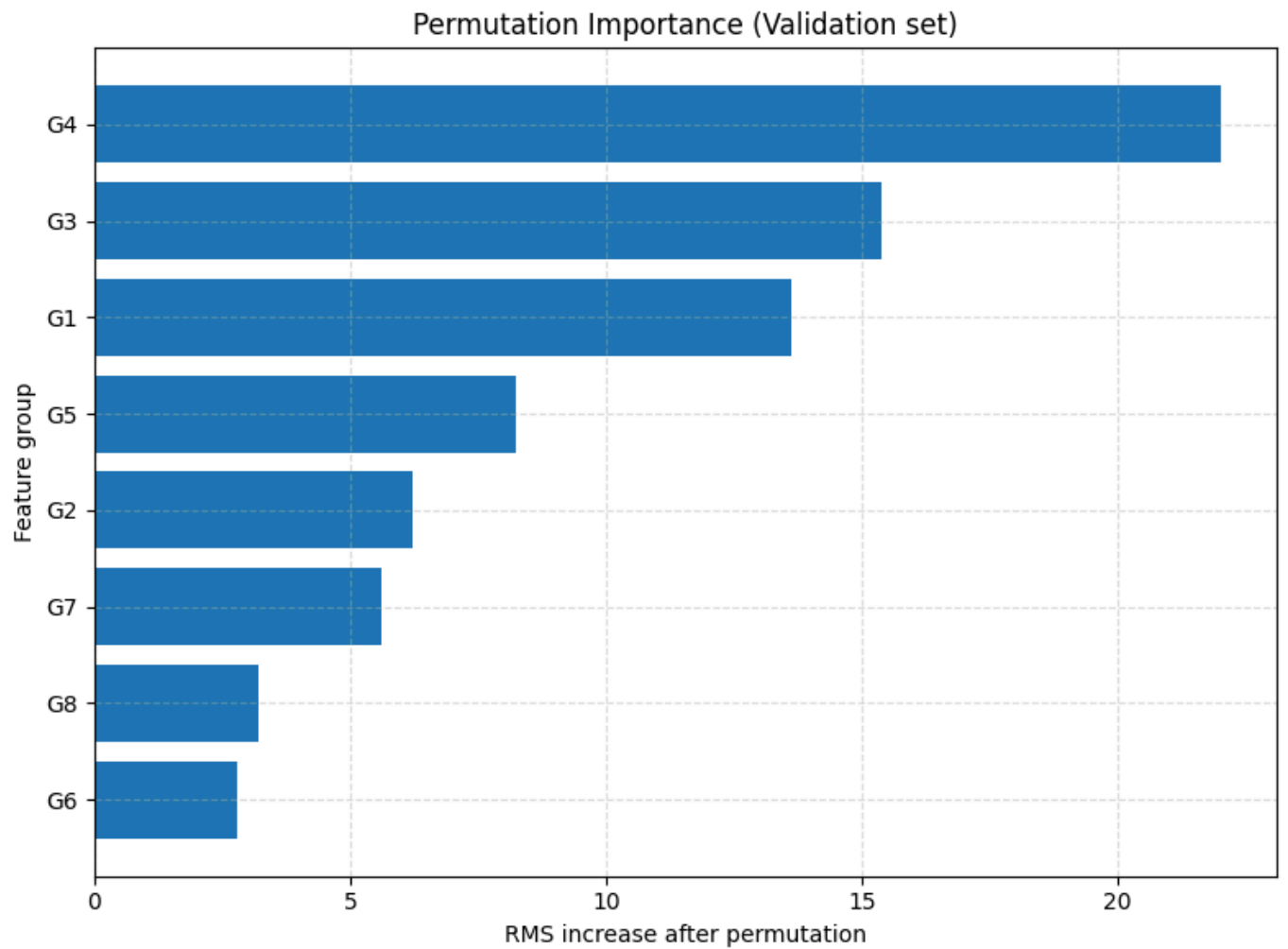
- 訓練/驗證/測試的散點緊貼對角線，可以看出模型有很好的捕捉到主要趨勢。
- 序列圖顯示模型能捕捉趨勢，但在極值樣本 (高負載的情況下) 有明顯低估，這可能是由於資料中高負載樣本較少，導致模型在這些區域的泛化能力較弱。
- 整體來說，模型在大部分樣本上表現良好，但在極端值上的表現仍有提升空間。

## 6. Feature Analysis 特徵分析

### 6.1 Permutation Importance

- 檔案：artifacts/regression/feature\_importance.csv ；視覺化：figures/regression\_feature\_importance.png 。
- 理論概念：**Permutation Importance** 以「擾動後效能下降量」衡量特徵重要性。假設模型在驗證集上的評分函數為  $\text{score}(X, y)$ ，當第  $j$  個特徵被隨機置換後，所得期望分數下降  $\Delta_j = \text{score}(X, y) - \text{score}(X_{\text{perm}(j)}, y)$ ，即為該特徵的貢獻度。由於本題使用 RMS 作為 score 的（分數越低越好），因此報告中的  $\Delta_{\text{RMS}} = \text{RMS}_{\text{perm}} - \text{RMS}_{\text{base}}$ ；值愈大表示該特徵在維持原始關聯性時提供愈多資訊。此方法不需重新訓練模型，能直接檢驗既有模型對特徵的依賴程度，也避免了梯度量測受尺度或分布影響的問題。
- 實驗方法：
  - 預先以早停訓練並保存最佳參數 ( best\_model.pkl )，評估時固定權重避免重新訓練。
  - 以驗證集作為評估基準，計算 baseline  $\text{RMS}_{\text{base}}=0.3760$ 。驗證集的使用可避免測試集資訊洩漏。
  - 先將特徵分組 ( g1 ~ g8 )，單一連續欄位為一組，one-hot 類別向量為一組；這在程式中透過 group\_mapping 實作，目的是避免多欄 one-hot 被拆散後重要度被稀釋。
  - 對某一組所有欄位逐欄打亂 ( rng.permutation )，保持該欄邊際分布但破壞其與其他欄、目標的對應關係，再以原模型推論新的  $\text{RMS}_{\text{perm}}$ 。
  - $\Delta_{\text{RMS}} = \text{RMS}_{\text{perm}} - \text{RMS}_{\text{base}}$  作為該組的重要度，最後依  $\Delta_{\text{RMS}}$  降序排序並輸出 CSV 與長條圖。
- Top-6 重要度 (RMS 上升量)：





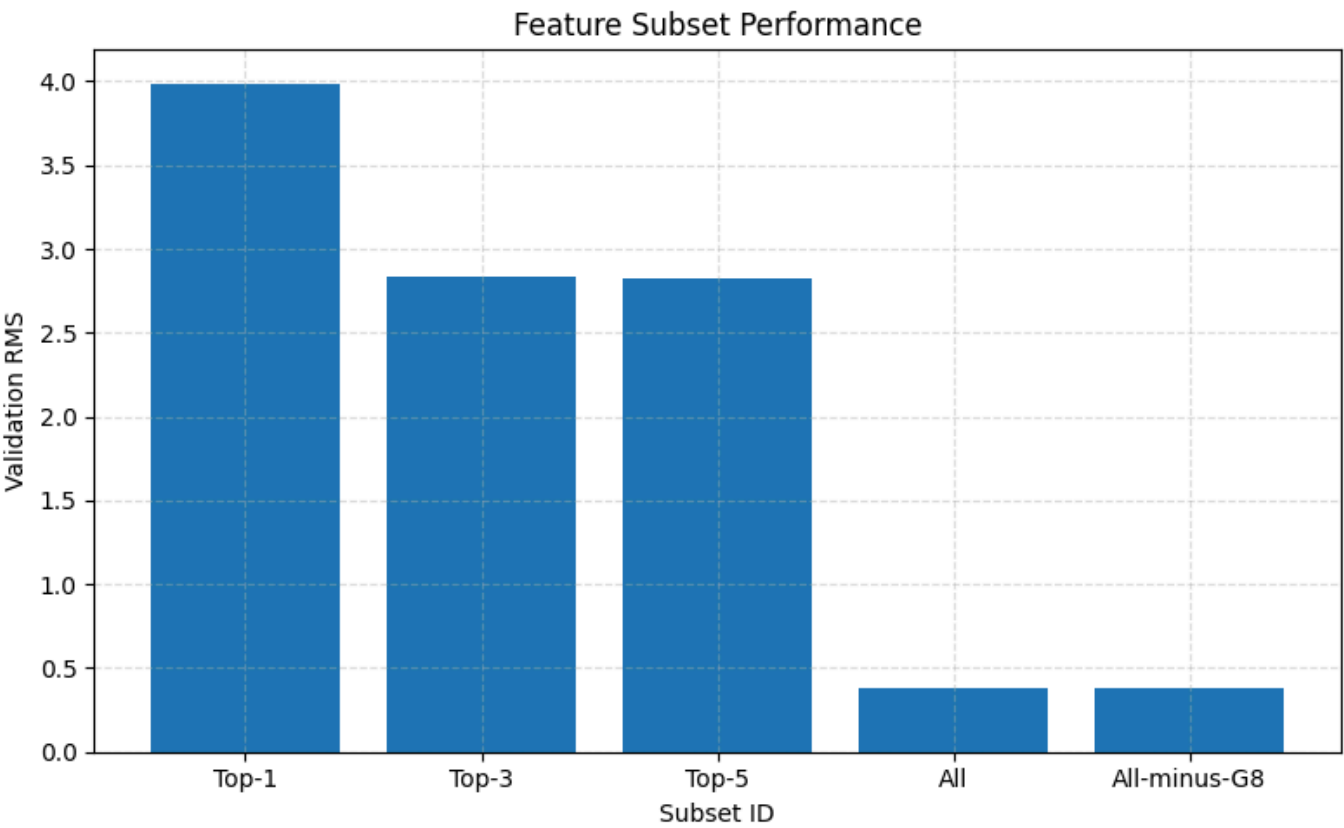
Rank	Group	Features	$\Delta$ RMS
1	G4	roof_area	<b>22.01</b>
2	G3	wall_area	15.40
3	G1	relative_compactness	13.63
4	G5	overall_height	8.23
5	G2	surface_area	6.23
6	G7	orientation_* (4 dims)	5.63
7	G8	gad_* (6 dims)	3.20
8	G6	glazing_area	2.78

- 觀察 Observations：
  - 面積相關特徵 ( roof\_area , wall\_area , relative\_compactness ) 重要度最高，顯示建築物的尺寸與形狀是影響暖氣負載的主要因素。
  - overall\_height 亦具顯著影響，可能因為高度影響熱量分布與流動。

- orientation (窗戶朝向) 也有中等影響，符合直覺。
- glazing\_area (窗戶面積) 重要度最低，可能因為資料中窗戶面積變化較小，或其影響被其他面積特徵所掩蓋。
- 整體來看，特徵重要度排序基本上與直覺相符，驗證了模型在學習過程中捕捉到合理的物理關聯。

## 6.2 Feature Subset Evaluation

- 檔案：artifacts/regression/feature\_subset\_results.csv ；圖片：figures/regression\_subset\_performance.png ◦
- 實驗方法 Experiment Setup：
  - 依照 Permutation Importance 排序，從最重要特徵開始逐步加入 (Top-1, Top-3, Top-5)，其他特徵設為 0。
  - 每組特徵皆重新訓練模型，並以相同超參數與早停條件。
  - 最終比較Top-K、全特徵與移除 gad\_\* (G8, 10 維) 的結果並以RMS進行排序。



Subset	Included Groups	#Features	Val RMS	Test RMS
Top-1	[ 'G4' ]	1	3.9880	4.3813
Top-3	[ 'G4', 'G3', 'G1' ]	3	2.8376	3.2808
Top-5	[ 'G4', 'G3', 'G1', 'G5', 'G2' ]	5	2.8285	3.2618

Subset	Included Groups	#Features	Val RMS	Test RMS
All	所有群組 (16 特徵)	16	<b>0.3760</b>	<b>0.4598</b>
All-minus-G8	排除 gad_*	10	0.3837	0.6706

- 觀察 Observations：
  - 僅使用頂尖特徵時 (Top-1~5) 誤差遠高於全特徵，顯示小型網路仍需多種特徵的輸入才能捕捉複雜度。
  - 移除 glazing\_area\_distribution ( G8 ) 造成測試 RMS 由 0.46 惡化至 0.67，說明窗戶分布雖未列入 Top-5，仍提供泛化所需訊息。

## 7. Appendix 附錄

- 主要檔案：
  - 程式：project/src/regression\_pipeline.py
  - 設定：project/src/regression\_pipeline.py::CONFIG
  - 模型權重：project/artifacts/regression/best\_model.pkl
  - 指標：project/results/regression\_summary.json
  - 圖表：project/figures/regression\_\*.png
- 重現步驟 Reproducibility：

```
cd HW1/project
python src/regression_pipeline.py
```

會自動完成前處理、訓練、評估與特徵分析；若要重新產出報告圖表請確保 figures/ 、 artifacts/ 可寫入。